Bootstrap based uncertainty bands for prediction in functional kriging

Maria Franco-Villoria*and Rosaria Ignaccolo

Department of Economics and Statistics "Cognetti de Martiis"

University of Torino

Abstract

The increasing interest in spatially correlated functional data has led to the development of appropriate geostatistical techniques that allow to predict a curve at an unmonitored location using a functional kriging with external drift model that takes into account the effect of exogenous variables (either scalar or functional). Nevertheless uncertainty evaluation for functional spatial prediction remains an open issue. We propose a semi-parametric bootstrap for spatially correlated functional data that allows to evaluate the uncertainty of a predicted curve, ensuring that the spatial dependence structure is maintained in the bootstrap samples. The performance of the proposed methodology is assessed via a simulation study. Moreover, the approach is illustrated on a well known data set of Canadian temperature and on a real data set of PM_{10} concentration in the Piemonte region, Italy. Based on the results it can be concluded that the method is computationally feasible and suitable for quantifying the uncertainty around a predicted curve.

Keywords: B-splines; band depth; functional data modelling; generalized additive models; geostatistics; trace-variogram

1 Introduction

Kriging is a well known prediction method in the geostatistics community (see e.g. [4]); it allows to predict a (scalar) random field or spatial process $\{Z(s), s \in D \subseteq \mathbb{R}^2\}$ in a new spatial location s_0 given a set of observed values $Z = (Z(s_1), \ldots, Z(s_n))$, taking into account the underlying correlation structure. Spatially dependent functional data (see e.g. the last two chapters of the book by [16]) have received increasing interest over the last few years. Geostatistical techniques for functional data were first introduced in the pioneering work of [15], but the development of such techniques is rather recent. The simplest case would be that of ordinary kriging, which

*Corresponding author. Email: maria.francovilloria@unito.it and rosaria.ignaccolo@unito.it

allows to predict a curve at an unmonitored site under the assumption of a constant mean (see e.g. [6, 13, 23]). The case of a mean function that depends on longitude and latitude was considered in [1, 22, 27]. In their work, [18] consider more complex forms of non-stationarity, where the mean function may depend on exogenous variables (either scalar or functional), developing the so called kriging with external drift - or regression kriging - in a functional data setting. While much effort has been put in prediction, the uncertainty of a predicted curve remains an open issue, since there is no functional version of the kriging variance. The lack of a distribution function in the functional framework leads to the use of resampling methods for confidence band calculation. In this context, [5] consider the standard bootstrap and a smoothed version of it to obtain confidence intervals for location estimators; an informal discussion on the asymptotic validity of the bootstrap approach in a functional framework can also be found in their paper. Further, [7] propose using "wild bootstrapping" in the case of a regression model with scalar response and functional covariate. When the response variable and hence the errors are both functional, to evaluate uncertainty of a predicted curve (obtained from kriging with external drift), we propose to extend the semi-parametric bootstrap approaches for spatially correlated data introduced by [30] and [19] to the case of functional data. The idea underpinning both works is to remove the spatial dependence structure, resample uncorrelated data and re-introduce the spatial correlation on bootstrapped samples. [19] consider the presence of a non-constant mean structure too, but the main difference between the two proposals is related to the considered statistics: [19] suggest to create the bootstrap distribution of the spatial predictor, while [30] construct a bootstrap distribution for the contrast defined as the difference between the spatial predictor and the unknown value (the one we want to predict).

The paper is organized as follows. In Section 2, we summarize the kriging with external drift (FKED) methodology developed in [18] and extend it to take into account spatial correlation when estimating the drift functional coefficients by means of an iterative algorithm. In Section 3 we propose a method for deriving prediction bands in the general FKED setting; in particular, we follow [30] and modify their approach to take into account a functional drift as proposed by [19]. A simulation study is presented in Section 4 to evaluate the performance of the proposed method, followed by an application to two real data sets. All computations are coded in R [24]. A discussion completes the paper.

2 Functional kriging with external drift (FKED)

Let $\Upsilon_s = \{Y_s(t); t \in T\}$ be a functional random variable observed at location $s \in D \subseteq \mathbb{R}^d$, whose realization is a function of $t \in T$, where T is a compact subset of $\mathbb{R}$. Assume that we observe a sample of curves Υ_{s_i}, for $s_i \in D$, $i = 1, \ldots, n$, taking values in a separable Hilbert space

of square integrable functions. The set $\{\Upsilon_s, s \in D\}$ constitutes a functional random field or a *spatial functional process* [6] that is not necessarily stationary. The following model is assumed:

$$\Upsilon_s = \mu_s + \epsilon_s, \tag{1}$$

where μ_s is the drift describing a spatial trend and ϵ_s is a zero-mean, second-order stationary and isotropic residual random field. At a fixed site s_i, $i = 1, \ldots, n$, and domain point t, the model can be rewritten as a functional concurrent linear model [18]

$$Y_{s_i}(t) = \mu_{s_i}(t) + \epsilon_{s_i}(t) \tag{2}$$

where $\epsilon_{s_i}(t)$ represents the residual spatial functional process $\{\epsilon_s(t), t \in T, s \in D\}$ at site s_i. The drift term can be expressed in terms of a set of scalar and functional covariates:

$$\mu_{s_i}(t) = \alpha(t) + \sum_p \gamma_p(t) C_{p,i} + \sum_q \beta_q(t) X_{q,i}(t) \tag{3}$$

where $\alpha(t)$ is a functional intercept, $C_{p,i}$ is the p^{th} scalar covariate at site s_i, $X_{q,i}$ is the q^{th} functional covariate at site s_i and $\gamma_p(t)$ and $\beta_q(t)$ are the covariate functional coefficients. Model (3) parameters can be estimated by means of a generalized additive model (GAM) representation using the R package `mgcv` (see [18], [34] and [36] for details).

To take into account the spatial correlation between functional observations when estimating the drift term, we propose an iterative algorithm that considers the term $\epsilon_{s_i}(t)$ as a functional random intercept (i.e. a location specific smooth residual) with a given covariance structure [29]. An iterative algorithm is also proposed in [22] to estimate drift coefficients for scalar covariates in universal kriging, as well as in [17] to take into account the heteroskedasticity of functional residuals. The algorithm can be summarized as follows:

1. Fit a standard functional concurrent linear model; estimate the drift term $\mu_{s_i}(t)$ following Model (3) assuming independent functional observations and obtain the functional residuals $e_{s_i}(t) = Y_{s_i}(t) - \hat{\mu}_{s_i}(t)$.

2. Estimate the correlation matrix K of the residual spatial functional process using the trace-semivariogram [13]. This is defined, for a zero-mean weakly-stationary isotropic process, as $\upsilon(h) = \int_T \frac{1}{2} Var\left(\epsilon_{s_i}(t) - \epsilon_{s_j}(t)\right) dt$, where $h = ||s_i - s_j||$ represents the Euclidean distante between locations s_i and s_j. The trace-semivariogram can be estimated as:

$$\hat{\upsilon}(h) = \frac{1}{2\,|N(h)|} \sum_{i,j \in N(h)} \int_T \left(e_{s_i}(t) - e_{s_j}(t)\right)^2 dt$$

where $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\}$. The estimate becomes computationally efficient when data are expressed using cubic B-splines, as integration can be avoided by re-expressing the integral in terms of the spline coefficients and basis [13]. Once estimated, the empirical trace-semivariogram provides a cloud of points $(h_g, \hat{v}(h_g)), g = 1, \ldots, G$ to which a parametric model (e.g. exponential, spherical, Matérn) can be fitted as in classical geostatistics.

3. Fit Model (2) considering the term $\epsilon_{s_i}(t)$ as a functional random effect, following [29], where the inverse of the estimated correlation matrix $\hat{K}$ is used as the precision matrix of a random field across locations.

4. Obtain the functional residuals $e_{s_i}(t) = Y_{s_i}(t) - \hat{\mu}_{s_i}(t)$. Return to step (2) and iterate until convergence.

The generalized additive model representation of Model (2) can be re-expressed as a mixed effects model [28, 31] whose parameters are estimated using REstricted Maximum Likelihood (REML) [35]. The algorithm's convergence is determined based on the Akaike Information Criterion AIC, since the effective degrees of freedom may change from iteration to iteration. The algorithm stops when the AIC rate is smaller than 0.1%, where the criterion rate at the j^{th} iteration is calculated as

$$AICrate = \left| \frac{AIC^j - AIC^{j-1}}{AIC^{j-1}} \right|.$$

The resulting functional residuals (at the last iteration) $e_{s_i}(t) = Y_{s_i}(t) - \hat{\mu}_{s_i}(t)$ can be used to predict the residual curve at an unmonitored site s_0 via one of three kriging options: 1) ordinary kriging for functional data [13], according to which $\hat{e}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i e_{s_i}(t)$, with kriging coefficients $\lambda_i \in \mathbb{R}$; 2) continuous time-varying kriging [12], where the kriging coefficients $\lambda_i(t)$ now depend on t and 3) functional kriging total model [11, 23], where the kriging coefficients are defined on $T \times T$ and

$$\hat{e}_{s_0}(t) = \sum_{i=1}^{n} \int_T \lambda_i(\tau, t) e_{s_i}(\tau) d\tau.$$

Prediction at the unmonitored site s_0 is obtained by adding up, as in the classical regression kriging, the two terms, i.e. $\hat{Y}_{s_0}(t) = \hat{\mu}_{s_0}(t) + \hat{e}_{s_0}(t)$, where $\hat{\mu}_{s_0}(t) = \hat{\alpha}(t) + \sum_p \hat{\gamma}_p(t) C_{p,0} + \sum_q \hat{\beta}_q(t) X_{q,0}(t)$ depends on the covariate values $C_{p,0}$ and $X_{q,0}(\cdot)$ at site s_0.

From now on we focus on the ordinary kriging case; this not only has been shown to be preferable when applied to a real dataset [18] but contributes to keep a moderate computational complexity in what follows (the two other cases will be considered in the discussion). Indeed, the ordinary case turns to be computationally convenient because of the trace-variogram's use. Not only integration can be avoided when using B-splines, but also it is possible to show (see appendix)

that the trace-variogram induces a covariance structure that is separable with respect to space and the domain of the functional data (e.g. time or depth). Moreover, [21] show that the solution to the Ordinary Kriging Problem via the trace-covariogram turns out to be the best finite-dimensional approximation of the operatorial kriging predictor for Hilbert-space valued random fields.

Note that, in practice, data are gathered as a finite discrete set of observations (t_j, y_{ij}), $t_j \in T$, $j = 1, \ldots, M$, $i = 1, \ldots, n$. Thus, before fitting Model (2), raw data need to be transformed into functional observations assuming $y_{ij} = Y_{s_i}(t_j) + \delta_{ij}$, where δ_{ij} represents measurement error and $Y_{s_i}(\cdot)$ is a continuous function that corresponds to a realization of the functional random field $\{\Upsilon_s, s \in D\}$ at site s_i. The conversion from discrete data to curves involves smoothing; we use cubic B-splines and the R package `fda` [26], choosing the number of basis functions and penalty parameter using functional cross-validation [18].

3 Bootstrap uncertainty bands for functional kriging

To evaluate the uncertainty of a predicted curve $\hat{Y}_{s_0}(t)$ at a new site s_0, we propose a semi-parametic bootstrap approach for spatially correlated functional data that builds on the work done by [30] and [19] for scalar data. The main idea is to decorrelate the data so that resampling can be done on independent observations and then transform back, ensuring that the spatial dependence structure is maintained in the bootstrapped samples. Previously we had only considered the extension of [19]'s work (see [10] for details), but simulation results were not stable (in terms of coverage). Now, after removing the drift as in [19], we propose to follow [30] according to which in the following algorithm a sample of size $n+1$ is drawn and an augmented covariance matrix created, such that a bootstrap datum $Y^{*j}_{s_0}$ is generated at the unmonitored location s_0.

Suppose that $\hat{Y}_{s_0}(t) - Y_{s_0}(t)$ follows the distribution F_n, a $1-\alpha$ prediction interval for $Y_{s_0}(t)$ can be built as $(\hat{Y}_{s_0}(t) - q_{1-\alpha/2}, \hat{Y}_{s_0}(t) - q_{\alpha/2})$, with q_α the α^{th} quantile of the unknown distribution F_n. The idea is to construct B bootstrap replicates $\{\hat{Y}^{*j}_{s_0}, Y^{*j}_{s_0}\}_{j=1}^B$ and approximate F_n by $\hat{F}^*_n$, the empirical distribution of $\{\hat{Y}^{*j}_{s_0} - Y^{*j}_{s_0}\}_{j=1}^B$. The bootstrapping algorithm can be summarized as follows:

1. Iteratively estimate the drift following Model (3) as proposed in steps 1-3 of the FKED algorithm to take into account the spatial correlation and take the functional residuals $e_{s_i}(t) = Y_{s_i}(t) - \hat{\mu}_{s_i}(t)$.

2. Estimate the functional residuals covariance matrix Σ through the estimated trace-semivariogram. Using Cholesky decomposition, $\hat{\Sigma} = \hat{L}\hat{L}^T$ and the functional residuals can be transformed

so that they become (spatially) uncorrelated:

$$\zeta_{n\times M} = (\zeta(s_1), \ldots, \zeta(s_n))' = \hat{L}^{-1}_{n\times n} \left(Y_{n\times M} - \hat{\mu}_{n\times M}\right).$$

3. Generate B bootstrap samples with size $n+1$, $\zeta^*_{n+1} = (\zeta^*(s_1), \ldots, \zeta^*(s_n), \zeta^*(s_{n+1}))'$ from $\zeta(s_1), \ldots, \zeta(s_n)$.

4. Create the augmented covariance matrix $\hat{\Lambda} = \begin{bmatrix} \hat{\Sigma} & \hat{c}_n^T \\ \hat{c}_n & \hat{\sigma}^2 \end{bmatrix}$, where $\hat{c}_n = \{\hat{C}(s_i - s_0)\}_{i=1}^n$, $\hat{C}$ is the estimated covariance function and $\hat{\sigma}^2 = \hat{C}(0)$ is the estimated scale. Use Cholesky decomposition so that $\hat{\Lambda} = \hat{R}\hat{R}^T$ and transform the bootstrap samples ζ^*_{n+1} as

$$(e^*(s_1), \ldots, e^*(s_n), e^*(s_0))' = \hat{R}_{(n+1)\times(n+1)} \zeta^*_{(n+1)\times M}.$$

5. The final bootstrap sample is determined as $Y^*_{s_i}(t) = \hat{\mu}_{s_i}(t) + e^*_{s_i}(t)$, $i = 1, \ldots, n$ and $Y^*_{s_0}(t) = \hat{\mu}_{s_0}(t) + e^*_{s_0}(t)$.

The bootstrap samples $\{Y^{*j}_{s_1}, \ldots, Y^{*j}_{s_n}\}_{j=1}^B$ are then fed into the FKED model to obtain B prediction curves $\hat{Y}^{*j}_{s_0}$ and the differences $\{\hat{Y}^{*j}_{s_0} - Y^{*j}_{s_0}\}_{j=1}^B$ are considered. The prediction interval for $Y_{s_0}(t)$ can be written as $(\hat{Y}_{s_0}(t) - q^*_{1-\alpha/2}, \hat{Y}_{s_0}(t) - q^*_{\alpha/2})$, with q^*_α the α^{th} percentile of $\hat{F}^*_n$, that can be obtained ordering the set of curves $\{\hat{Y}^{*j}_{s_0} - Y^{*j}_{s_0}\}_{j=1}^B$. However, the idea of ordering curves is not as straightforward as ordering scalar values, and to our knowledge there is no gold standard for doing so. We consider two different ordering techniques available in the literature.

The first one builds on the idea of band depth [20], that can be defined for any set of k curves. In their paper, [20] suggest using k=3, stating that there is no need to increase k as "the band depth induced order is very stable in k". Even for k=2, the computational cost is considerable when the sample includes a large number of functional curves; to improve computation times, we have adopted the fast algorithm proposed in [32] and [33] with k=2. The band in $\mathbb{R}^2$ delimited by the curves y_{i_1}, y_{i_2} is defined as

$$B(y_{i_1}, y_{i_2}) = \left\{(t, y(t)) : t \in T, min_{r=1,2}\, y_{i_r}(t) \leq y(t) \leq max_{r=1,2}\, y_{i_r}(t)\right\}.$$

The sample band depth (BD) of a curve $y(t)$ in a set of n curves can be calculated as

$$BD_{n,2}(y) = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{1\leq i_1 < i_2 \leq n} I\left\{G(y) \subseteq B(y_{i_1}, y_{i_2})\right\}$$

where $G(y)$ is the graph of a curve $y(t)$ defined as the subset of the plane $G(y) = \{(t, y(t)) : t \in I\}$.

Potential problems of using $k = 2$ include ties (i.e. more than one curve with the same depth value) and crossing over of the curves delimiting the band (in which case the band "is degenerated in a point and, with probability one, no other curve will be inside this band"; see [20]). To avoid these problems and still count with the computational advantage of using $k = 2$, band depth can be modified to take into account whether a portion of the curve is in the band, giving rise to the modified band depth (MBD), defined as

$$MBD_{n,2}(y) = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \frac{\lambda\left(\{t \in T : min_{r=i_1,i_2} y_r(t) \leq y(t) \leq max_{r=i_1,i_2} y_r(t)\}\right)}{\lambda\left(T\right)}$$

where λ is the Lebesgue measure on T (for further details see [20]). With this scheme, the bigger the band depth value, the more central the curve is.

The second ordering scheme is based on L^2 distance between curves [5]. In this case, the bootstrap-based predicted curves are ordered based on how distant they are from the zero curve, according to the L^2 distance definition:

$$||x - y|| = \left(\int_T \left(x(t) - y(t)\right)^2 dt \right)^{1/2}. \tag{4}$$

With this scheme, the smaller the distance, the more central the curve is.

The lower/upper limits of a 95% prediction band are obtained by taking the pointwise (w.r.t. t) minimum/maximum of the 95% deepest curves (using band depth) (i.e. those closest to the center of the distribution) [32] or of the 95% curves closest to the zero curve (using L^2 distance) [5].

4 Simulation Study

The performance of the bootstrapping method proposed in Section 3 is evaluated here through a simulation study. Our aim is to analyse the impact of the spatial structure of the functional residual random field, by means of the covariance function parameters (scale and range), as well as that of the ordering technique chosen to derive the uncertainty bands when increasing the number of sites.

Data were simulated using cubic B-splines on a spatial irregular grid (n locations) on $D = [0, 2] \times [0, 3]$ and curve domain $T = [0, 1]$. We used a B-spline basis on T with 10 basis functions. The residual functional random field was built as $e_s(t) = \sum_{j=1}^{10} \xi_j(s) B_j(t)$, where $B_j(t)$ is the j^{th} basis function evaluated at $t \in T$ (i.e. a curve) and $\{\xi_j(s), s \in D\}$ are the spatially correlated spline coefficients. These were generated for each $j = 1, \ldots, 10$ using the same exponential

covariance function with range and scale parameters $\phi \in (0.5, 1, 1.5)$ and $\sigma^2 \in (0.25, 0.50, 0.75)$ respectively, resulting in 9 different scenarios. As already mentioned in Section 2 and shown in the appendix, the trace-variogram of $\{e_s(\cdot), s \in D\}$ turns out to be proportional to the variogram of the spline coefficients $\{\xi_j(s), s \in D\}$. The drift was obtained as

$$m_s(t) = \alpha(t) + \beta_1(t)lon + \beta_2(t)lat$$

where *lon* and *lat* are the spatial coordinates, $\alpha(t)$ is a functional intercept and $\beta_1(t), \beta_2(t)$ are functional coefficients. The functional coefficients $\alpha(t)$, $\beta_1(t)$ and $\beta_2(t)$ can be expressed in terms of the same B-spline basis with scalar spline coefficients (common for all sites) that are drawn from normal distributions as follows: for $\alpha(t)$ we draw 10 splines coefficients from $N_{10}(\mathbf{1}, 0.05\, I_{10\times 10})$; for $\beta_1(t)$ and $\beta_2(t)$ we draw 10 splines coefficients from $N_{10}(\boldsymbol{\vartheta}, 0.05\, I_{10\times 10})$ with $\boldsymbol{\vartheta} = (0.2, 0.2, 0.4, 0.4, 0.6, 0.6, 0.8, 0.8, 1, 1)^T$, where $I_{10\times 10}$ represents the identity matrix with dimension 10×10. Finally, simulated observations were built as

$$Y_s^{sim}(t) = m_s(t) + e_s(t) + \eta_s(t)$$

where $\boldsymbol{\eta}(t) = \{\eta_{s_1}(t), \ldots, \eta_{s_n}(t)\} \sim N_n(\mathbf{0}, 0.09\, I_{n\times n})$ is a vector of random errors for each fixed $t \in [0, 1]$; in practice we consider 101 equally spaced points in $[0, 1]$.

For each simulation scenario, we generated functional data at $n = 25, 50$ and 90 nested locations. Additionally, data were generated at 10 more sites (always the same for all three sample sizes) used as validation stations. The locations can be seen in Figure 1, while the simulated data can be seen in Figure 2 for $n = 90$ (note that the cases $n = 25$ and $n = 50$ are just subsets of this).

For each of the three sample sizes, and each simulation scenario, the FKED model presented in Section 2 was applied to the corresponding data set to predict curves at the 10 validation sites. In particular, a drift term depending on longitude and latitude was considered and ordinary kriging was used to obtain the predicted residuals. In practice, the variogram model is chosen automatically among exponential, gaussian and spherical based on minimum SSE. Computational times ranged from 2 seconds (n=25) up to 16 seconds (n=100). The resulting predicted curves, along with the observed data, can be seen in Figure 3 in the case of $n = 90$ and range and scale parameters $\phi = 0.5$ and σ^2=0.25, respectively. The estimated range and scale parameters $\hat{\phi}$ and $\hat{\sigma}^2$ for all 9 simulation scenarios are in agreement with what one would expect given the values set for ϕ and σ^2 in the simulation design and the relationship between the trace-variogram and the variogram described in the appendix. It appears that there is good accordance between simulated and predicted observations. Similar figures for the remaining cases are available upon request as supplementary material. Following the algorithm illustrated in Section 3, a bootstrap

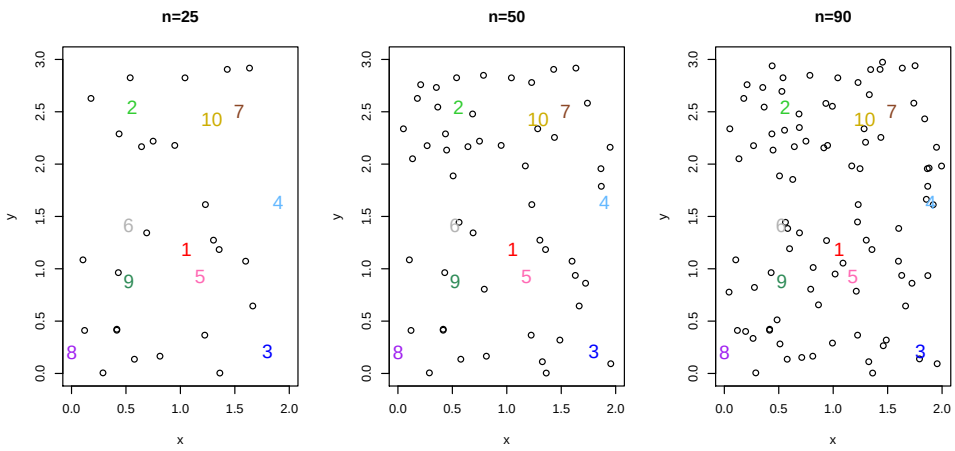

Figure 1: Locations of the 25, 50 and 90 sites used for model fitting. Validation sites numbered 1 to 10.

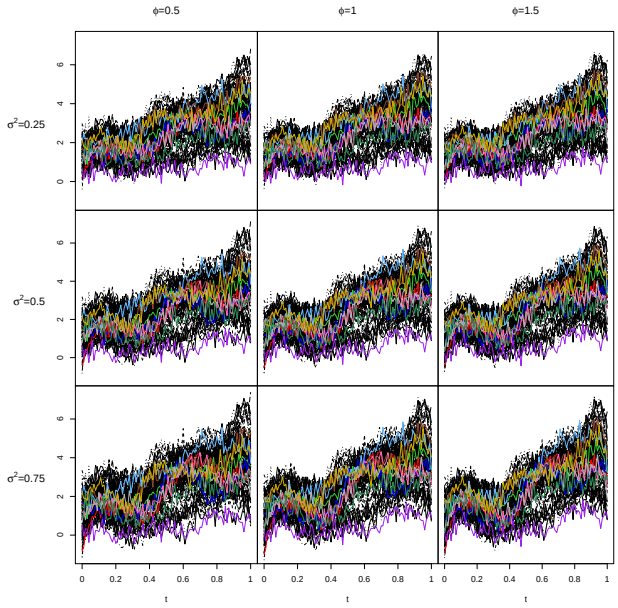

Figure 2: Simulated data (n=90) with 10 validation stations curves in color, for each simulation scenario.

sample of size $B = 500$ was obtained for each validation station. These 500 curves were ordered using both distance and band depth, where the latter was calculated using the modified version *MBD*. The resulting 95% prediction bands are shown in Figure 3. Overall, the two uncertainty measures provide very similar prediction bands.

To evaluate the performance of the bootstrap method proposed, we can use different indicators. First of all, we consider the width of the resulting 95% prediction interval. Then we

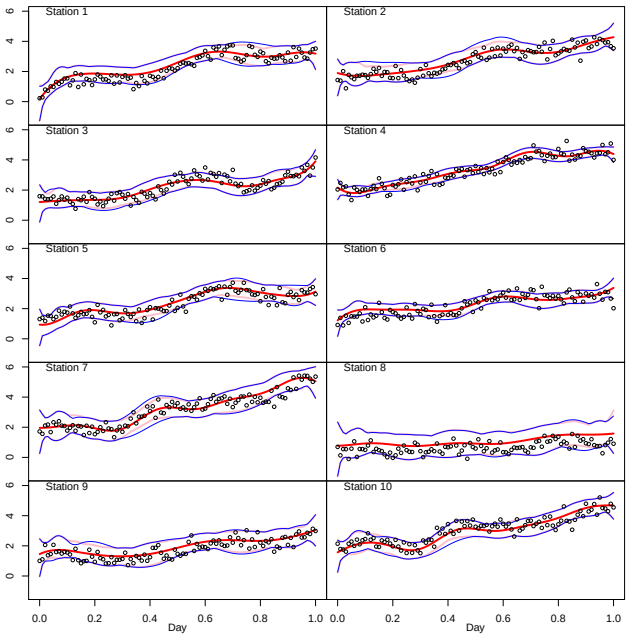

Figure 3: Original data (black dots), FKED predicted curve (solid red line), 95% prediction band based on L^2 distance (pink) and on MBD (blue) for n=90, $\sigma^2 = 0.25$, $\phi = 0.5$.

evaluate the "domain coverage" that is the proportion over the domain T of the simulated curve within the prediction band. The latter should not be understood as coverage in the classical sense, and thus we propose a "functional coverage" defined as the percentage of times that the prediction band contains the true curve in a set of S simulations; here $S = 100$.

In Figure 4 (left), we show the depth based band width corresponding to a sample size of $n = 25$ and all simulation scenarios, while Figure 4 (right) summarizes the difference in width when using band depth and distance for the same sample size. The remaining figures (for $n = 50, 90$) are not shown here due to space limitations but are available upon request as supplementary material. As expected, the width of the interval in the curve domain becomes greater towards the borders. As for the effect of the covariance function parameters, greater values of the scale parameter σ^2 lead to wider prediction bands, while for a fixed value of the scale parameter, the width of the band decreases slightly with increasing range ϕ. When comparing band depth and distance (Figure 4 (right)) it is interesting to see that the depth based interval is predominantly wider than the distance based one, regardless of the value of σ^2 and ϕ. Figure 5 shows the depth based prediction band width for a fixed simulation scenario ($\sigma^2 = 0.5$ and $\phi = 1.5$) and all three samples sizes; here it can be seen that as n increases, the width of the interval decreases and it becomes slightly more stable on T.

On the other hand, the domain coverage is summarized in Figure 6 for all simulation scenarios

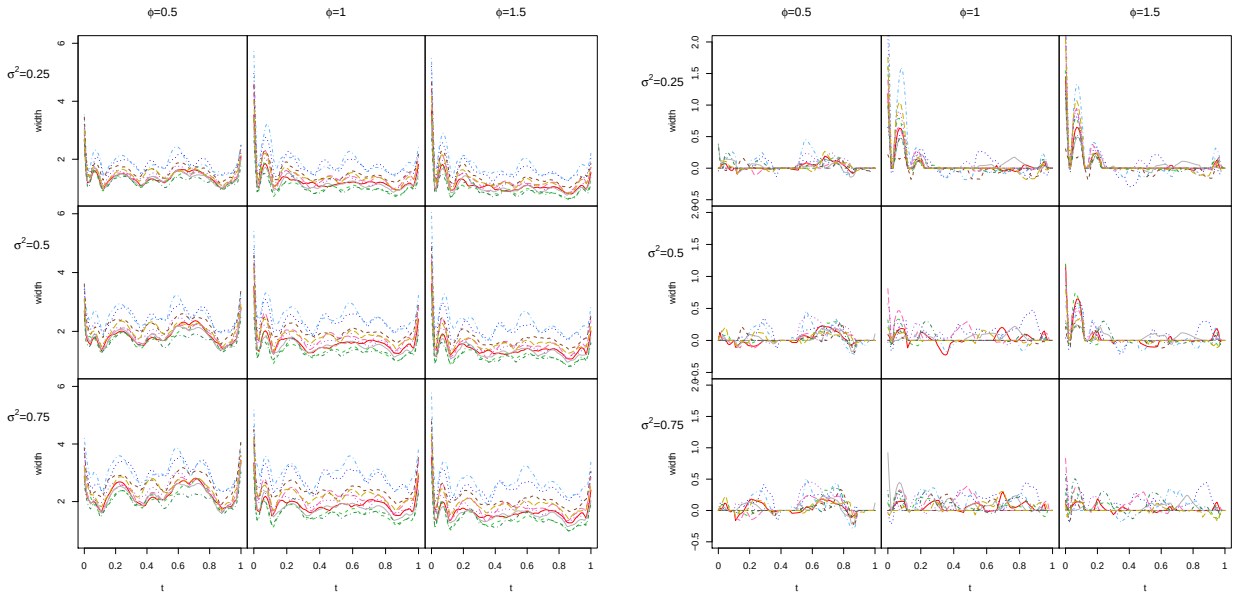

Figure 4: Prediction band width according to band depth (left) and width (depth-distance) difference (right) for 10 validation stations and different simulation scenarios when $n = 25$.

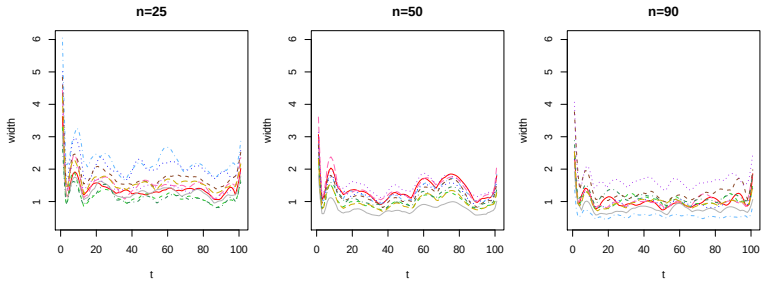

Figure 5: Prediction band width, when using band depth for ordering curves, for a fixed simulation scenario ($\sigma^2 = 0.5$ and $\phi = 1.5$) and all three samples sizes.

and sample sizes. This coverage is linked to the width of the interval (the larger the width the larger the coverage) but also to the goodness of FKED prediction because the uncertainty band is built around the predicted curve: if the prediction is far from the original data, the corresponding coverage will be poor, and viceversa. We can see in Figure 6 that the domain coverage improves as n increases and it is slightly better for small values of σ^2. Overall, coverage ranges from 87.1% to 100% and is highly dependent on validation site. In particular, it can be seen that when $n = 90$ all predicted curves apart from those corresponding to validation stations numbered 4 and 7 are fully contained in the prediction band. For $n = 25$ and $n = 50$ the number of validation stations with coverage smaller than 100% increases due to the fact that fewer locations lead to poorer predictions, but the domain coverage in these cases is always greater than 87.1%.

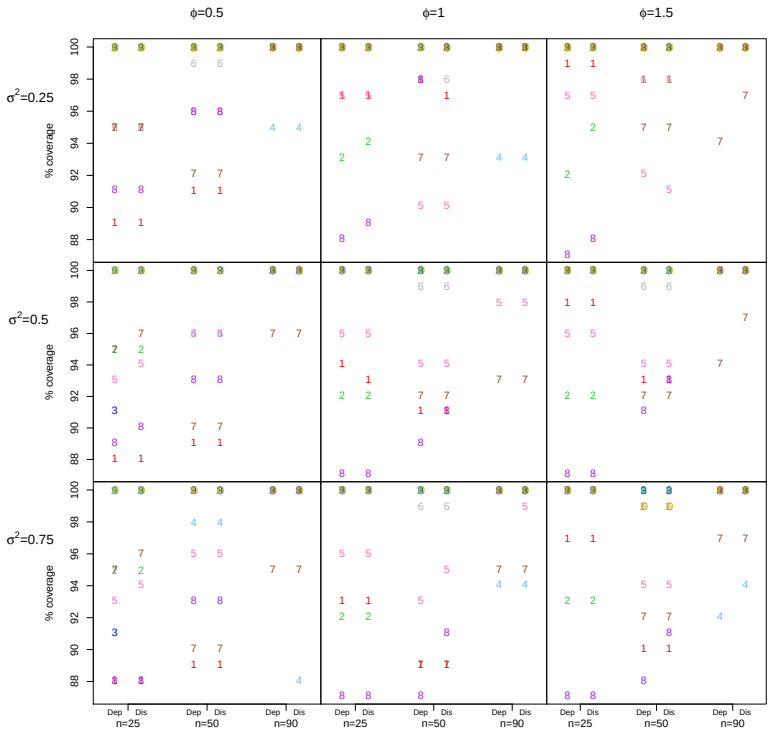

Figure 6: Prediction band coverage for n=25, 50 and 90 and all 9 simulation scenarios.

Regarding functional coverage, it decreases as σ^2 increases but improves with increasing values of ϕ (i.e. when the dependence structure becomes stronger). If we constrain the whole true curve to be contained in the prediction band - that is fixing the domain coverage equal to 100% - then the functional coverage ranges from as little as nearly 0 in the worst scenario (small sample size, high variability, weak spatial dependence) up to 100%, as shown in Figure 7 (left). Even though this might look like poor performance in the first instance, we must bear in mind that the functional coverage varies greatly with validation site. Once again, sites 4 and 7 can be identified as poor in terms of performance for $n = 90$; if we disregard these two, the coverage for the remaining stations ranges between 80% and 100% in the best scenario ($\sigma^2 = 0.25$, $\phi = 1.5$). Further, if we allow for some tolerance on the domain coverage, e.g. a coverage of at least 90%, functional coverage improves considerably as shown in Figure 7 (right), even for those sites that seemed to perform poorly.

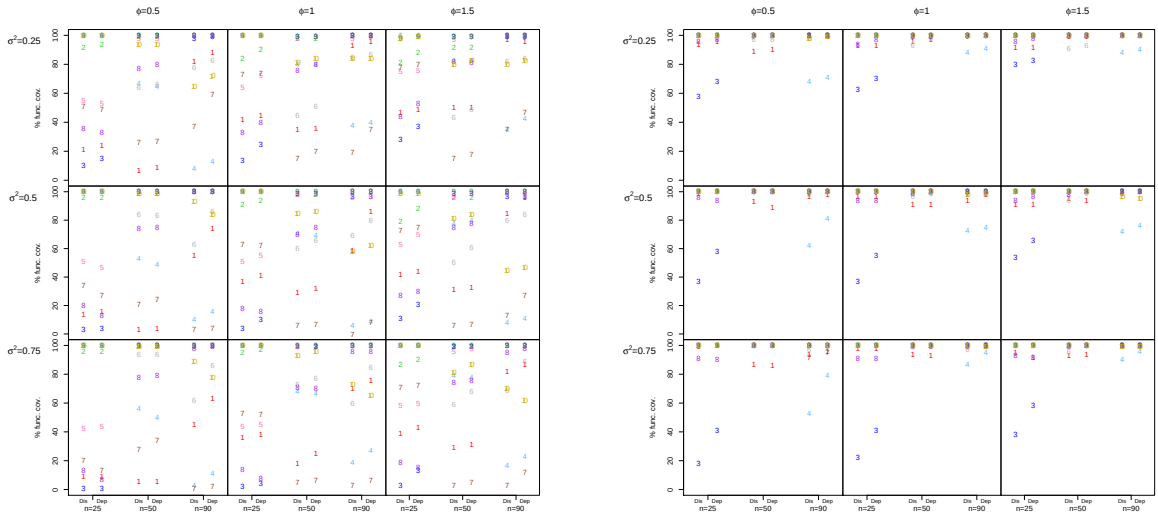

Figure 7: Functional coverage for n=25, 50 and 90 and all 9 simulation scenarios, with domain coverage equal to 100% (left) and larger than 90% (right).

5 Real data analysis

5.1 Canadian temperature data

As first case study, we have chosen the well known data set of Canadian temperature data. This has been repeatedly used in the functional data literature (see, for example, [22, 12, 25, 29]). The data set consists of daily annual mean temperature collected at 35 meteorological stations in Canada's Maritimes Provinces: Nova Scotia, New Brunswick and Prince Edward Island, over the period 1960 to 1994. Note that the data set used here (available in the `geofd R` package; [14]), which is the same as in [22, 12], covers a smaller geographical area than that in [25]. While other papers have been devoted to prediction of these temperature curves, our objective is to provide uncertainty bands for the predicted curves, and hence discussing the fitted model is out of the scope of this paper. To this aim, we have selected five stations at random to use as validation stations for which we will provide prediction bands according to our proposal in Section 3. These can be seen in red in Figure 8 (left). Data were converted to functional observations through smoothing by using penalized cubic B-splines with 120 basis functions and penalty parameter equal to zero. These values were chosen using functional cross-validation. The FKED model, with longitude and latitude as covariates, was then fitted to the remaining 30 stations and predicted temperature curves were obtained (using ordinary functional kriging with an exponential variogram model) for the 5 validation stations. From the empirical trace-variogram, there was no evidence of a discontinuity at the origin and hence we fixed the nugget equal to zero. For each of the validation sites, a bootstrap sample of size $B = 1000$

was obtained. Band depth was calculated using the modified version *MBD*. The resulting 95% prediction bands are shown in Figure 9. The uncertainty bands are fairly narrow, as expected when observing the small variability among curves in Figure 8 (right) but they become slightly wider in winter. Overall, the two uncertainty measures seem to agree well, although in some cases the distance based prediction band appears to be slightly narrower than the depth based one. Domain coverage percentages for all 5 validation sites range from 98.9% to 100%.

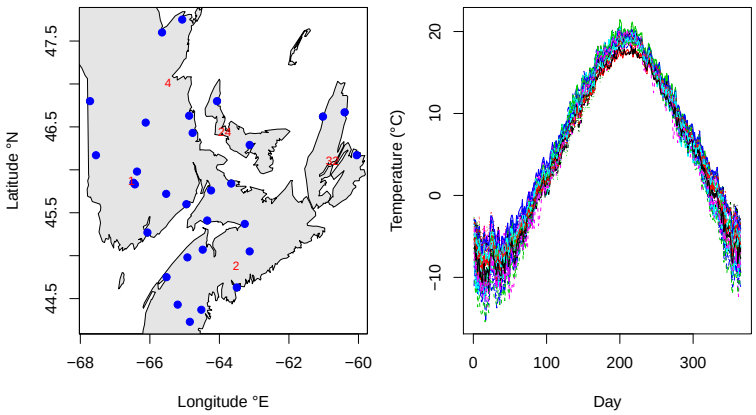

Figure 8: Locations of the 35 meteorological stations in Canada's Maritimes Provinces area (left, validation stations numbered in red) and temperature curves (right, raw data)

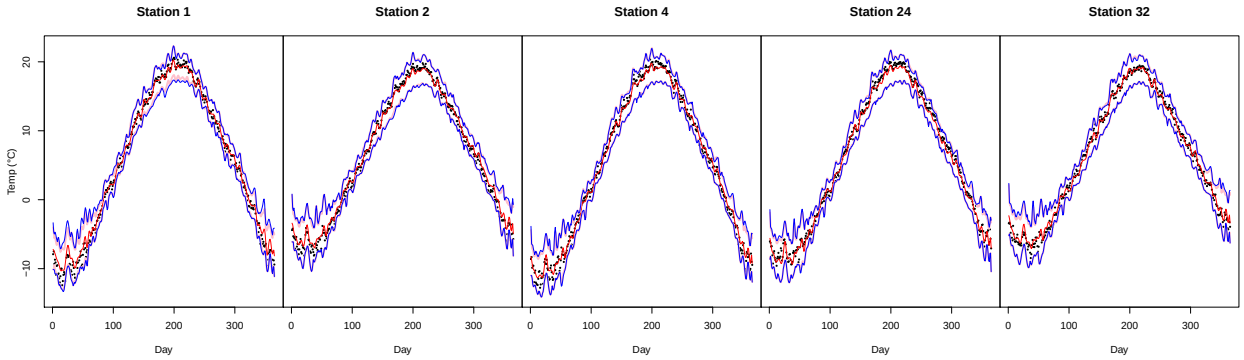

Figure 9: Original temperature data (black dots), FKED predicted curve (red line), 95% prediction band based on L^2 distance (pink) and on *MBD* (blue) for validation stations

5.2 Air pollution data

The second case study considered consists of daily PM_{10} concentrations (in $\mu g/m^3$) measured in 24 sites (red triangles in Figure 10) from October 2005 to March 2006 by the monitoring network

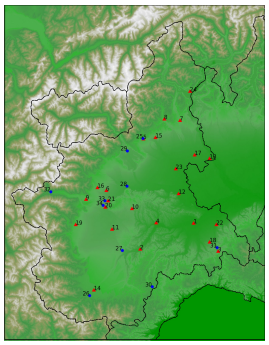

Figure 10: Locations of the 24 PM_{10} monitoring sites (red triangles) and 10 validation stations (blue dots).

of Piemonte region (Italy). Measurements were also available at 10 additional locations (blue dots in Figure 10) that are used as validation stations. Apart from geographical information, i.e. longitude, latitude and altitude of each station, considered as scalar covariates, information was available on daily maximum mixing height, daily total precipitation, daily mean wind speed, daily mean temperature and daily emission rates of primary aerosols, taken as functional covariates. These are available thanks to a nested system of deterministic computer-based models implemented by the environmental agency ARPA Piemonte [8]. This data set had already been analyzed in [2]. For further details on the model the reader is referred to [18]. A log transformation was used on the response variable to achieve normality and stabilize within-station variability. Prior to modelling, data (both response and functional covariates) were smoothed by means of cubic B-splines with 146 basis functions and penalty parameter equal to 0. These values were chosen using functional cross-validation [18].

The functional kriging with external drift model described in Section 2 was applied to the air pollution data, including the (standardized) scalar and functional covariates mentioned above, to obtain prediction curves (via ordinary kriging for functional data with an exponential variogram model and zero nugget) at the 10 validation sites. To obtain 95% prediction bands for each of the predicted curves, a bootstrap sample of size 1000 was obtained for each site following the algorithm proposed in Section 3. We obtained prediction bands according to both the modified band depth (with $k = 2$) and distance induced order. Prediction bands for the ten validation sites are shown in Figure 11. Overall, the two uncertainty measures seem to agree well, although in some cases the depth based band appears to be slightly wider than the distance based one. The domain coverage varies from 97.3% to 100%.

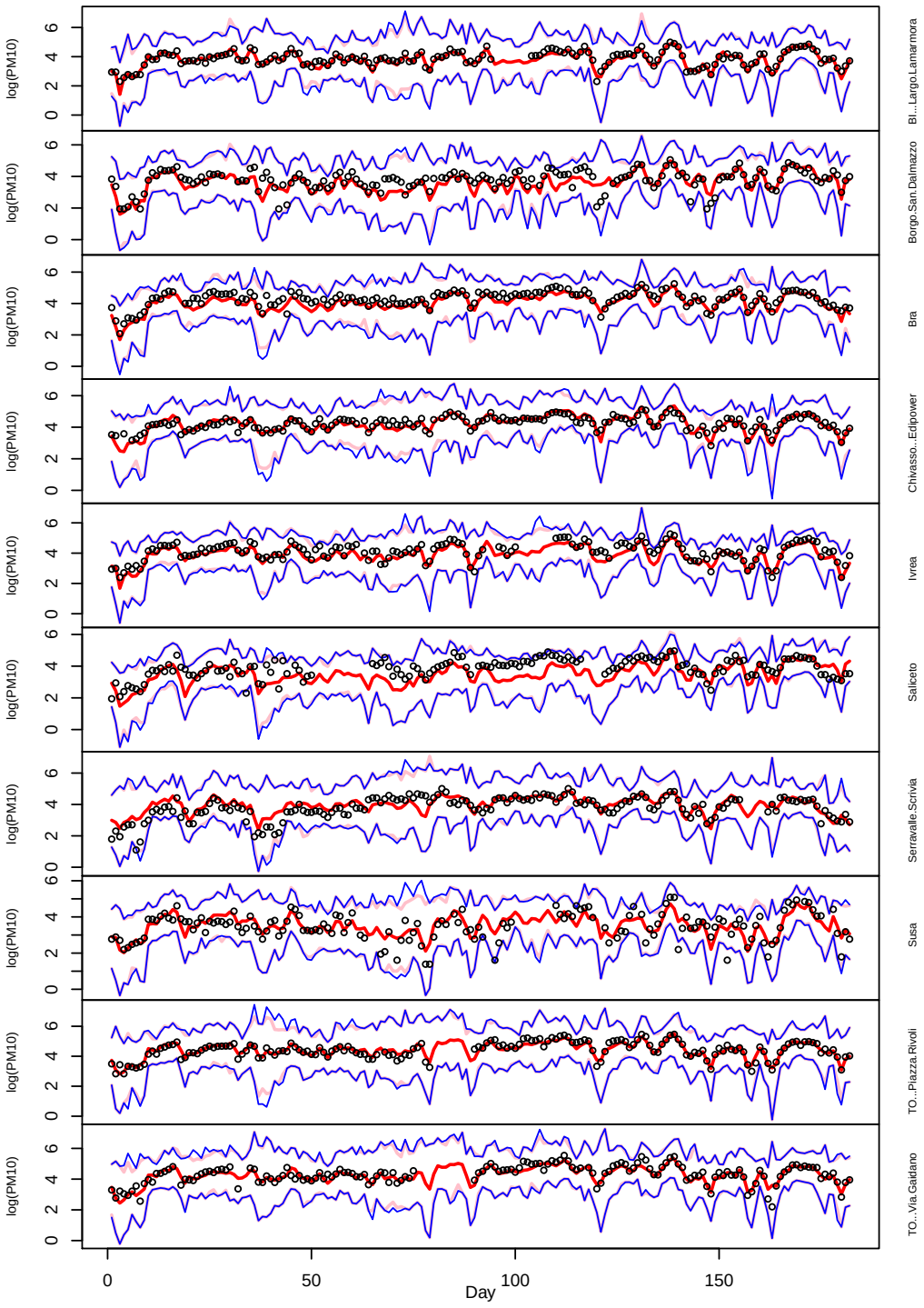

Figure 11: Original PM_{10} data (black dots), FKED predicted curve (red line), 95% prediction band based on L^2 distance (pink) and on *MBD* (blue) for validation stations.

6 Discussion

The functional kriging with external drift model allows to predict a whole curve - regardless of the domain of the functional observations - taking into account exogenous covariates and the underlying spatial dependence. Nevertheless, uncertainty evaluation in the kriging context has not been addressed so far. The classic functional kriging variance provides a unique value over the whole domain T, while our aim is to provide an uncertainty measure whose value may change along the predicted curve. Given the lack of an analytic expression of the kriging variance for a curve, we propose a semiparametric bootstrap approach that not only allows the uncertainty to change over the domain T but also takes into account the uncertainty due to drift estimation. The contrast considered to build the uncertainty band, defined as the difference between the spatial prediction and the unknown value, could be thought of as a sort of pivotal quantity, the distribution of which is unknown. The use of a pivotal quantity is recalled to improve the performance of the bootstrap in the parametric framework [3].

We considered two different techniques proposed in the literature for ordering the bootstrapped curves, namely functional depth and L^2 distance, but we did not find great differences in the results for either the simulations or the two case studies. To be conservative, one could argue that it is safer to use the band depth as this criterion gives larger bands on average.

Overall, using the known data at validation sites in both simulations and real case studies, it can be concluded that the proposed method appears to be a valid approach in evaluating uncertainty for a predicted curve. In particular, the simulation study shows that the width of the bands depends on the scale and range parameters of the covariance structure, at least for the exponential case, and becomes more stable over the domain T with increasing sample size.

In a previous work [9] for the PM_{10} case study, we also considered, with a longitudinal point of view, a generalized additive model (GAM) with a smooth function of longitude, latitude and time to predict a curve, so that standard inference results apply for deriving uncertainty bands. In the considered case study, FKED prediction outperformed GAM prediction even though with a larger uncertainty; nevertheless, the prediction interval provided by the GAM model is pointwise (and it could be corrected e.g. by the Bonferroni method) while our approach gives a simultaneous prediction band w.r.t. T.

Data in the two case studies considered in this paper were regularly spaced over the domain T; however, that is not always the case (see e.g. atmospheric profiles in [17]). For irregularly spaced data, curves could be aligned at the initial smoothing step before fitting the functional kriging with external drift model, as it is straightforward to evaluate the curves for every $t \in T$ in a common grid for all curves.

The algorithm proposed for uncertainty evaluation of spatially correlated curves is computa-

tionally feasible (running times with $B = 1000$: 4 hours for the PM_{10} case study and 2.35 hours for the Canadian temperature case study using an Intel Core i7-4770 CPU 3.40GHz 16GB RAM) and applies to a wide range of practical situations, as it can be used regardless of the complexity of the drift term, or even in the absence of the latter. Our proposal has been specified in the case of ordinary kriging where the weight coeficients are constant and the spatial structure is determined by means of the trace-variogram. We are aware that the simulation study is limited to the simple case in which the spatial dependence structure does not change w.r.t. $t \in T$, as assumed by the trace-covariogram. While it might be worth exploring more complex scenarios a full analysis of the air pollution data described in Section 5.2 [18] suggested that in practice the assumption of a spatial structure that is constant over time is reasonable, as the drift term in the model picks up most of the spatial variability over time.

Nevertheless, it may be the case that a more complex kriging alternative is desiderable in order to let the weights vary with t (for both the real data and the bootstrap samples). To be consistent with the way the kriging weights have been specified in the cases of continuous time-varying kriging and functional kriging total model, the underlying spatial structure (and hence the matrices K in Section 2 and Σ in Section 3) should not be determined by means of the trace-variogram. In fact, in [13] and [11] functional data are expressed as linear combinations of splines and a Linear Model of Coregionalization is used to estimate cross-correlations among the spline coefficients; this way a possible interaction between the curve domain and the space domain is taken into account. Consequently the matrices K and Σ would need to be adjusted accordingly but with the added burden of an increased computational load.

We believe that the method proposed in this paper is appropriate in the framework of functional data and is able to provide uncertainty bands for a predicted curve in an unmonitored site. Further, half the width of the resulting prediction band could be considered as an approximate margin of error. We think this will prove useful for monitoring purposes and policy assessment where the uncertainty should always accompany the related prediction.

7 Acknowledgments

This work was supported by "Futuro in Ricerca" 2012 Grant (project no. RBFR12URQJ, named StEPhI) provided by the Italian Ministry of Education, Universities and Research. The authors are grateful to Alan Gelfand and all the participants of the StEPhI intermediate workshop for the fruitful discussion.

References

[1] W. Caballero, R. Giraldo, and J. Mateu. A universal kriging approach for spatial functional data. *Stochastic Environmental Research and Risk Assessment*, 27(7):1553–1563, 2013.

[2] M. Cameletti, R. Ignaccolo, and S. Bande. Comparing spatio-temporal models for particulate matter in piemonte. *Environmetrics*, 22(8):985–996, 2011.

[3] A.J. Canty, A.C. Davison, D.V. Hinkley, and V. Ventura. Bootstrap diagnostics and remedies. *Canadian Journal of Statistics*, 34(1):5–27, 2006.

[4] J.-P.l Chiles and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty, 2nd Edition.* Wiley, Hoboken, NJ, 2012.

[5] A. Cuevas, M. Febrero, and R. Fraiman. On the use of the bootstrap for estimating functions with functional data. *Computational Statistics and Data Analysis*, 51:1063–1074, 2006.

[6] P. Delicado, R. Giraldo, C. Comas, and J. Mateu. Statistics for spatial functional data: some recent contributions. *Environmentrics*, 21:224–239, 2010.

[7] F. Ferraty, I. Van Keilegom, and P. Vieu. On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics*, 37:286–306, 2010.

[8] S. Finardi, R. DeMaria, A. D'Allura, C. Cascone, G. Calori, and F. Lollobrigida. A deterministic air quality forecasting system for Torino urban area, Italy. *Environmental Modelling and Software*, 23(3):344–355, 2008.

[9] M. Franco-Villoria and R. Ignaccolo. Kriging uncertainty for functional data: a comparison study. In *Proceedings of the METMAVII–GRASPA 14 Conference. Special issue of GRASPA Working Paper*, 2014.

[10] M. Franco-Villoria and R. Ignaccolo. Uncertainty evaluation in functional kriging with external drift. In E.G. Bongiorno, E. Salinelli, A. Goia, and p.Vieu, editors, *Contributions in infinite-dimensional statistics and related topics (Proceedings of the Third International Workshop on Functional and Operational Statistics (IWFOS)*, pages 113–118. Societ Editrice Esculapio, 2014.

[11] R. Giraldo, P. Delicado, and J. Mateu. Geostatistics with infinite dimensional data: a generalization of cokriging and multivariable spatial prediction. Technical report, Reporte Interno de Investigacion No. 14, Universidad Nacional de Colombia, 2009.

[12] R. Giraldo, P. Delicado, and J. Mateu. Continuous time-varying kriging for spatial prediction of functional data: an environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):66–82, 2010.

[13] R. Giraldo, P. Delicado, and J. Mateu. Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, 18(3):411–426, 2011.

[14] R. Giraldo, J. Mateu, and P. Delicado. geofd: An r package for function-valued geostatistical prediction. *Revista Colombiana de Estadstica*, 35(3):383–405, 2012.

[15] M. Goulard and M. Voltz. Geostatistical interpolation of curves: A case study in soil science. In A. Soares, editor, *Geostatistics Troia 92*, pages 805–816. Kluwer Academic, Dordrecht, 1993.

[16] L. Horváth and P. Kokoszka. *Inference for functional data with applications*. Springer, New York, 2012.

[17] R. Ignaccolo, M. Franco-Villoria, and A. Fassò. Modelling collocation uncertainty of 3D atmospheric profiles. *Stochastic Environmental Research and Risk Assessment*, 29(2):417–429, 2015.

[18] R. Ignaccolo, J. Mateu, and R. Giraldo. Kriging with external drift for functional data for air quality monitoring. *Stochastic Environmental Research and Risk Assessment*, 28:1171–1186, 2014.

[19] N. Iranpanah, M. Mohammadzadeh, and C.C. Taylor. A comparison of block and semi-parametric bootstrap methods for variance estimation in spatial statistics. *Computational Statistics and Data Analysis*, 55:578–587, 2011.

[20] S. Lopez-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American Statistcal Association*, 104(486):718–734, 2009.

[21] A. Menafoglio and G. Petris. Kriging for Hilbert-space valued random fields: The operatorial point of view. *Journal of Multivariate Analysis*, 146:84–94, 2016.

[22] A. Menafoglio, P. Secchi, and M. Dalla Rosa. A universal kriging predictor for spatially dependent functional data of a Hilbert space. *Electronic Journal of Statistics*, 7:2209–2240, 2013.

[23] D. Nerini, P. Monestiez, and C. Manté. Cokriging for spatial functional data. *Journal of Multivariate Analysis*, 101:409–418, 2010.

[24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[25] J. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, New York, 2006.

[26] J. O. Ramsay, Hadley Wickham, Spencer Graves, and Giles Hooker. *fda: Functional Data Analysis*, 2014. R package version 2.4.4.

[27] A. Reyes, R. Giraldo, and J. Mateu. Residual kriging for functional spatial prediction of salinity curves. *Communications in Statistics - Theory and Methods*, 44(4):798–809, 2015.

[28] G.K. Robinson. That BLUP is a good thing: the estimation of random effects. *Statistical Science*, 6:15–32, 1991.

[29] Fabian Scheipl, Ana-Maria Staicu, and Sonja Greven. Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2):477–501, 2015.

[30] L. Schelin and S. Sjöstedt de Luna. Kriging prediction intervals based on semiparametric bootstrap. *Math. Geosci.*, 42:985–1000, 2010.

[31] T. Speed. Comment on paper by Robinson. *Statistical Science*, 6:421744, 1991.

[32] Y. Sun and M. Genton. Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334, 2011.

[33] Y. Sun, M. Genton, and D.W. Nychka. Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat*, 1:68–74, 2012.

[34] SN. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2006.

[35] SN. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36, 2011.

[36] SN. Wood. *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*, 2015. R package version 1.8.6.

Appendix: note on the trace-variogram

Let us recall that the trace-variogram is defined, for a zero-mean weakly-stationary isotropic process, as

$$\upsilon(h) = \int_T \frac{1}{2} Var\left(\epsilon_{s_i}(t) - \epsilon_{s_j}(t)\right) dt$$

where $h = ||s_i - s_j||$ represents the Euclidean distante between locations s_i and s_j.
By assuming a finite development for the functions $\epsilon(t)$ such that $\epsilon_{s_i}(t) = \sum_{l=1}^{Nb} \xi_l(s) B_l(t)$ where Nb is the number of considered basis functions and $B_l(t)$ is the l-th basis function evaluated at $t \in T$, we can write

$$Cov\left(\epsilon_{s_i}(t), \epsilon_{s_j}(v)\right) = \sum_{l=1}^{Nb} Cov\left(\xi_l(s_i), \xi_l(s_j)\right) B_l(t) B_l(v) = C(s_i, s_j) \sum_{l=1}^{Nb} B_l(t) B_l(v) = C(s_i, s_j)\tau(t, v)$$

where $C(s_i, s_j) = Cov\left(\xi_l(s_i), \xi_l(s_j)\right)$ is assumed identical for all l and $\tau(t, v) = \sum_{l=1}^{Nb} B_l(t) B_l(v)$.
Thus we get a factorization with respect to the spatial domain D and the curve domain T.
Moreover, for every i, $Var\left(\epsilon_{s_i}(t)\right) = C(s_i, s_i)\tau(t, t)$ and we can write

$$\begin{aligned} Var\left(\epsilon_{s_i}(t) - \epsilon_{s_j}(t)\right) &= Var\left(\epsilon_{s_i}(t)\right) + Var\left(\epsilon_{s_j}(t)\right) - 2Cov\left(\epsilon_{s_i}(t), \epsilon_{s_j}(t)\right) \\ &= C(s_i, s_i)\tau(t, t) + C(s_j, s_j)\tau(t, t) - 2C(s_i, s_j)\tau(t, t). \end{aligned}$$

With the assumption of stationarity and isotropy for ϵ we obtain

$$\upsilon(h) = [C(0) - C(h)] \int_T \tau(t, t) dt$$

so that the trace-variogram is written as a product of a (classical) spatial variogram and a constant depending on the Nb basis functions B_l.